MODIS
Distributed Active Archive Centers (DAACs)
Data Archive, Distribution, and User Services

Presentation to MODIS Science Team

January 24, 2001

Steve Kempler

301-614-5765

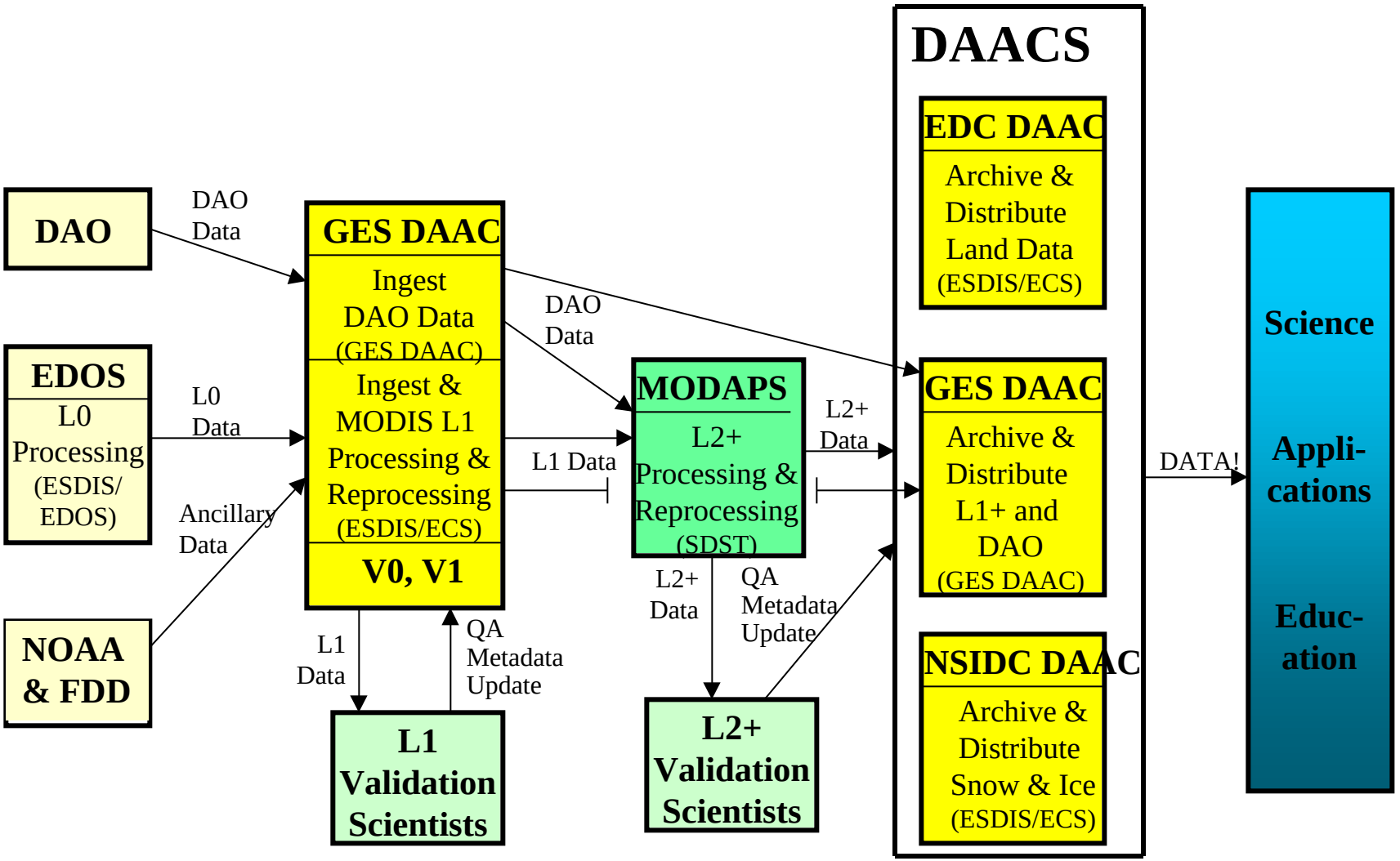
steven.kempler@gsfc.nasa.gov

With inputs from

Greg Scharfen (NSIDC DAAC) and John Dwyer (EDC DAAC)

MODIS Data Flow

(Development Team in parenthesis)



Archive Metrics at the MODIS DAACs

	TB/month sustained *	Total (TB), so far	Current Capacity (TB)
GSFC	17.4	135.4	750 **
NSIDC	0.31	2.5	60 **
EDC	3.5	29.5	250 **
TOTAL	21.2	167.4	1060 **

* Last three months averaged

** Assumes compression factor of 1.5:1

Archive Issues

- Everything is OK in the Archive!
- As DAACs become more stable, we need to take a good look at archive cleanup (i.e., delete unneeded and/or duplicate files, etc.)
- DAACs also will be analyzing what the results of data compression to determine the true compression factors
- This will lead to clear planning for additional archive to stay ahead of archive demand, and/or better define and implement a data product deletion/rotation plan

Distribution Volumes (average GB/Day) to Users

Year 2000	GSFC	NSIDC *	EDC
Feb	1.3		
Mar	6.5		
Apr	4.9		
May	10.9		
Jun	16.0		1.2
Jul	15.7		0.8
Aug	22.2		1.9
Sep	22.1		1.1
Oct	31.4	0.3	14.3
Nov	25.8	0.3	21.7
Dec	22.9	0.9	14.8
Distribution Capacities (GB/Day) theoretical:			
	~ 446	16	77
TOTAL DISTRIBUTED TO USERS, THUSFAR:			
	5.5 TB	**45 GB**	**1.7 TB**

***only one product suite (snow) available starting in October, test orders not included**

Is MODIS Data Distribution Lower Than it Should Be?

James Koziana, supported by A K Sharma,, queried participants at the recent AGU:

1. **Are you aware of the MODIS products that are available?**
2. **If the products are of interest to you, and you have not ordered any, what stopped you or why did you decide not to?**

- Less than 10% were not familiar with MODIS before visiting either the DAAC booth or poster presentations.
- Sample size: >350 people, from fields ranging from Education (K-12 and Introductory remote sensing at Colleges) to Atmospheric and Oceanographic Research (Model and Observational).

The reasons for not ordering MODIS data can be grouped into several categories.

1. **Data Maturity** - Several people indicated that they would wait for a more mature data product because of their resource limitations. *This is a typical occurrence for new data products*
2. **Data Access** - Several people had experienced difficulties with ordering data and have not tried again. *We need to let people know that data access has improved, and how to get help*
3. **Data Subsetting (spatial, temporal and parameter)** - The transfer of large data files is difficult. Several users would find the data more manageable if vertical profiles of selected parameters for specified geographical regions were available - *Excellent point. Many subsetting/data mining efforts going on.*
4. **Data Formats** - Some users expressed concern about the data only available in HDF-EOS format - *May examine the implementation of data format translators*

Distribution Issues

- Large orders must be handled manually:
  - Large orders have failed - fix recently installed at NSIDC
  - They are either worked off slowly (broken into smaller orders)
  - Or, if greater than distribution capacity, rejected
  - Analysis ongoing to determine true distribution capacity
  - Thus, maintaining a conservative approach to public outreach until products can be distributed with confidence
- Limited hard media distribution
  - EDC is developing the Product Distribution System (PDS) to be deployed by all DAACs
  - PDS will increase distribution capability of data via hard media (8 mm tape, digital linear tape, CD-ROM and DVD-R disks) without impacting ECS performance

The Earth System Information Partnership

- Context: The ESIP Federation.
  - Broadening of DIS activities for EOS.
    - Includes DAACs (ESIP 1s)
    - ESIP 2s (Science Data Centers) and
    - Application Centers (ESIP 3s).
- ESIPs have an understanding of requirements of particular communities and their needs.
- Decision at the Tucson Meeting of the Federation in early Jan 2001 to set up a MODIS Cluster

Provided by J. Townshend

MODIS Cluster

- **Objective of MODIS cluster**: Facilitate access to MODIS data
- **Strategy**: A Federation partnership between current MODIS data providers (ESIP 1s DAACs) and other ESIPs willing to contribute resources to help deliver MODIS data.
- **Capacity**: Help DAACs meet current *and future* distribution requirements
- **Packaging**: Provide alternative granularity and formats
- **Sources**: Provide better access to non-DAAC MODIS sources
  - Direct broadcast
  - PI processing

Provided by J. Townshend

MODIS Cluster:

- What Members Will Do
  - Replicate and distribute products
    - at least 1 TB online
    - access/ordering mechanism
  - Report metrics
  - Do share of cluster-wide work
- What The Cluster Will Do
  - Work collectively with data sources
    - DAACs
    - others…
  - Specify standard product-delivery metrics
  - Set up a “MODIS Clearinghouse”
    - Who has what for when/where

Provided by J. Townshend

Reprocessing Preparations Status

The GSFC DAAC is preparing for the first MODIS reprocessing campaign

- **Hardware**: 2X processing operational (1X: Processing forward; 1X reprocessing, then 2X (April) until Aqua comes on-line)
- **Software**: GES DAAC developed S4P will be operational to reprocess MODIS data in ECS, March
- **Level 0 Data**: Goal is to have all Level 0 data resident in DAAC. Complete: Apr/00, Sept.- Dec/00
- **Ancillary Data**: Goal is to have all ancillary data resident in DAAC. Complete: Apr/00, everything else fairly close
- **Science Software:** Must have new versions of science software 2 weeks prior to start of reprocessing, for integration and test
- **ESDTs**: Goal is to have all new ESDTs resident in DAAC. Status: Best approach to efficiently developing new ESDTs being analyzed by MODIS and DAAC personnel

User Services Experiences

- Number of User requests:
  - GSFC: ~5800 NSIDC: 86 EDC: ~1200
- GSFC
  - Requests for large data orders can not be filled
  - Requests for "real-time" data from the DAAC.
- NSIDC:
  - System limitations cause orders to fail or suspend:
    - DAAC staff troubleshoot problem, advise users
    - DAAC submits trouble tickets; ECS fix patches installed
- EDC
  - Order Shipping or FTP Notifications occasionally bounce back to User Services
  - Requests for large data orders can not be filled or broken into smaller orders
  - EDG is too complicated and confusing for some users
  - Searching on 0º longitude fails (latitude also?)

User Services Frequently Asked Questions

- How to register as a new users
- How to order a granule if I know the granule-id
- Can I arrange for subscription (ordering) services
- Questions regarding the data pickup procedures
- Questions regarding the EOSDIS Data Gateway (EDG)
- Questions concerning ordering problems
- Questions regarding browse and subsetted data
- How to use and work with HDF-EOS?
- How to geolocate MODIS data? (Why was ISIN grid chosen?)
- How to incorporate grids in data manipulation programs?
- What is the turn-around time for MODIS data availability
- (acquisition --> access)
- Where can I get information on the MODIS projection
- How does MODIS FTP work relative to that for AVHRR
- Couldn't retrieve FTP orders due to firewall on user end

MODIS Data Access

Ways to Access MODIS Data

- EOS Data Gateway (EDG)
  - Access directly or via DAAC Home Pages through Links button
  - Data is sent by tape or pulled by requestor
- DAAC-specific interfaces
  - GSFC: Terra-WHOM
- Anonymous FTP
  - Access via DAAC Home Page through On-line FTP button
  - Sample products first
  - Plans for rolling archive of samples later
- Subscription (currently for MODIS Science Team)
  - Data automatically pushed to requestor

DAAC User Services

- GSFC:
  - MODIS Data Support Team. Lead: Greg Leptoukh 301-614-5253
  - User Services Group:
    - 301-614- 5473 (ECS) or 301-614-5224 (V0)
    - daac_usg@gsfcsrvr4.gsfcmo.ecs.nasa.gov (ECS)
      daacuso@daac.gsfc.nasa.gov
    - http://daac.gsfc.nasa.gov
  - Special Data Processing Requests: Coordinate with discipline lead. Request goes to Ed Masuoka for cross discipline coordination.
- NSIDC
  - MODIS Snow and Ice Product Team. Lead: Greg Scharfen 303-492-6197
  - ECS Science Outreach Coordinator: Siri Jodha Singh Khalsa 303-492-1445
  - User Services:
    - HELP DESK - 303-492-6199
    - nsidc@kryos.colorado.edu
    - http://nsidc.org/NASA/MODIS/
- EDC DAAC
  - User Services MODIS Data Specialist: Carolyn Gacke 605 -594-6822
  - ECS MODIS Science Data Specialist: Calli Jenkerson 605 -594-2638
  - User Service:
    - Brenda Jones: 605 -594-6503; Toll Free 866 -573-3222 (866 LPE DAAC)
    - edc@eos.nasa.gov
    - http://edcdaac.usgs.gov

MODIS Data at the DAACs
Summary & Prognosis

- Baseline functionality delivered by ECS
  - Several tools built by DAACs to facilitate system operability
  - More fully being automated by ECS development and internal DAAC personnel
- Performance:
  - Processing and Archive appears adequate but still needs definitive testing (of new hardware upgrades)
  - True distribution capacities cautiously being tested and analyzed
- User Services personnel keeping up with increasing demand

- January 24, 2001 prognosis:
  - ECS performance requirements 'met' with theoretical performance capacities integrated. Real performance being analyzed by increasing loads.
  - Functionality will continue to become more reliable (automated) with ECS and DAAC enhancements
  - Ingest capacity to meet full MODIS product suite requirements need to be re-evaluated
  - Unprocessed data will be picked up with first reprocessing campaign; Reprocessing will no doubt, come with initial glitches
  - Need to prepare for Aqua data loads (i.e., Aqua capacities being made used now)

Level 0 is flowing; Unplanned downtime has decreased;
System confidence slowly building